VIET NAM LIVING STANDARDS SURVEY

Recommended Sample Design

Chris Scott, Consultant. November 1991

1. INTRODUCTION

This report results from detailed discussions held with officials of the State Planning Committee and the General Statistical Office under the guidance of Mr Tran Ngoc Trang, Director of the Institute of Planning Research of the SPC and Project Director, and of Mr Pham Van So, responsible for the Living Standards Survey.

I would like to thank the numerous participants in these meetings for their frank and effective collaboration. I also wish to thank those who have helped me at various times during my mission, notably Prof. Nguyen tu Qua, Mr Nguyen Dinh Ton and Mrs Ngo Thanh Hang, all from the SPC, Mr Walden of UNDP and Dr Macrae of UNFPA.

A field trip was made on 27 November to one urban and one rural commune. (See Annex for details.) We met with excellent cooperation from the local officials and I would like to thank Mr Hieu and Mr Du, the Head and Acting Head respectively of the two communes.

2. OBJECTIVES AND CONSTRAINTS

The purpose of this report is to describe an appropriate sample design for the Viet Nam Living Standards Survey to be undertaken in 1992. Such a design has to take account of the specific objectives of the survey, the resources available and the arrangements planned for the implementation of the survey in the field. In this section these aspects will be examined in more detail.

2.1 Sample size

Sample size is determined principally by the financial resources available, while taking account of the unit costs of the operations required. The detailed calculations are given in the UNDP Project Document. This provides for a sample of approximately 2500 households, based on a total project cost of $516,000.

However this figure would provide a barely adequate sample for

comprehensive analysis, leaving very little scope for regional disaggregation. Discussions are in course on the possibility of increasing the sample size by recourse to the Swedish consultancy fund under SIDA. An additional $200,000 is under discussion. This would allow an increase to 4,800 households.

At the time of writing this report no decision has yet been taken on whether these funds can be made available. The report therefore covers 2 options, one assuming a sample of 2,500 households, the other of 4,800.

2.2 Sample allocation between domains

The basic objectives assume that the analytic importance of any domain of study in the population should be roughly represented by the relative population size of the domain. (A region that is twice as populous should be considered twice as important to the analyst.) On this basis, and assuming that nothing is known about differential variances and operational costs between domains, the starting point for sample design is that the sample should be distributed in proportion to population, in other words that the same sampling fractions (or "sampling rates") should be applied throughout the country.

However one very important exception appears immediately. The urban sector in Viet Nam accounts for only about 20% of the population. In a sample distributed in proportion to population, a total sample of 2,500 households will yield only 500 in the urban sector, and a total of 4,800 will yield almost 1,000. Considering the wide variety of economic activities found among the urban population it does not seem reasonable to argue that a sample of 500 households would be adequate for analytic requirements.

In circumstances of this kind, where a particular domain is small in size but large in analytic interest, it is a common practice in many surveys to "over-sample" the domain of interest. This means applying an inflated sampling rate to that domain with the exclusive purpose of increasing the sample size for the domain. When this is done, a corrective weight must be applied whenever it is desired to put together data from the special domain with data from the rest of the sample. For example, if the urban sector is over-sampled by a factor of 2, then a weight of 1/2 would be applied to the urban data whenever they are combined with rural data.

In discussions with SPC and GSO it was agreed that this strategy should be applied in the event that the small sample has to be used (Option 1; 2,500 households). If funds are available for the large sample (Option 2), the total urban sample will reach about 1,000 and it seems unnecessary to over-sample the urban sector.

Thus, in Option 1, the sample will be distributed in proportion to population within the urban sector and also in proportion to population within the rural sector, but the sampling rate will be twice as high in the urban sector.

In Option 2, the sample will be in proportion to population throughout, with the same sampling rate everywhere.

The needed sampling rates will be calculated in Section 5.

2.3 Sampling units and sampling frames

If households were selected at random all over Viet Nam the distance between households in the sample would be very great and interviewers would spend nearly all their time traveling. For efficient field work it is necessary to concentrate the household sample in clusters. First a sample of areas is selected, then in each area a sample of households.

In this section area sampling is discussed first, then household sampling.

2.3.1 Area sampling

The structure of the official area division of Viet Nam is as follows:

Provinces	
RURAL	URBAN
Districts	Centers (= towns)
	Quarters (Quan)
Communes (Xa)	Communes (Phuong)
Villages (Thon)	Clusters (Cum)

NOTES

1. The number of provinces was 44 at the time of the census and is 49 today.
2. The rural population is about 80%; the urban about 20%.
3. The number of communes is about 10,000 and their average population about 6,500.
4. Villages and clusters have an average population somewhat less than 1,000.
5. There are smaller units within the urban clusters, called Groups, or To.

The villages and urban clusters would make very suitable area sampling units: their size is suitable and not too variable from one unit to another. Unfortunately their individual sizes, in terms of population or numbers of households, are not directly available to the SPC/GSO. Only the communes can provide such data. To obtain them from each commune in Viet Nam would be a huge operation.

On the other hand, the population and numbers of households totaled for each commune is available to SPC/GSO from the census records.

Thus a convenient solution would be a two-stage sample at the area level: communes at the 1st stage, villages or urban clusters at the 2nd stage. The procedure would be to select a sample of communes, then write to each commune office requesting a list of villages (rural case) or clusters (urban case), then select a sample of such units.

2.3.2 Household sampling

In each selected village or urban cluster a number of households must now be selected. This requires an up-to-date list of households from which to make the selection.

Once again, this information is available at the commune office. As soon as the villages/clusters have been selected, the SPC/GSO will write again to the commune office asking for a list of all households in those selected units.

One small problem remains: the definition of the household. In general there is no ambiguity in the concept of a household in Viet Nam: each household has to have a household ID booklet, showing each of the household members by name, which must be kept always up-to-date. The existence of this system ensures that the household as a unit is exactly defined. The only problem arises in relation to "collective households". These are groups of adult workers, all of the same sex and all belonging to the same enterprise, who are housed together although they are unrelated. Often they have families living elsewhere. The average size of such households is about the same as that of family households. Collective households can constitute about 10% of households in urban areas but are very rare in rural areas.

Careful discussion of this problem led to the conclusion that members of collective households would best be treated as one-person households. Each such individual is responsible for his own expenditures and one of them cannot respond on behalf of another. For example, each one can answer the question "How much do you send back to your family as remittances?", but he cannot answer this

question for another member.

Thus it will be necessary to make special arrangements for listing and sampling such cases in the urban sector. The commune must be asked to provide the number of persons living in collective households in each cluster (cum), at the 1st stage, and to include the list of such persons living in the selected cluster (cum) at the 2nd stage when asked for the list of households. For sampling purposes each such individual will be counted as a one-person household.

2.1 Sample allocation between sampling stages

The three sampling stages, as seen above, will be:

1st stage: Communes
2nd stage: Villages (rural) or Clusters (urban)
3rd stage: Households

This section asks how many households should be selected per village or cluster, how many villages/clusters should be selected per commune, and how many communes should be selected. The answers will be mainly determined by the arrangements adopted for field work. These arrangements will be first reviewed.

2.1.1 Field organization

Each household has to be interviewed twice. In the 2nd interview there are questions about expenditures during the 2-week interval since the interviewer's previous visit. It is therefore essential to allow a 2-week interval between the two visits.

The interviews are long and on the average an interviewer can only do 2 per day.

Each team has 2 interviewers.

Thus 16 households will be selected in each village/cluster and it is expected that the team will complete the 1st interviews for all 16 households in the 1st week. This village/cluster may be called Area Unit A.

At this point the completed questionnaires are taken to the data entry operator (located in a nearby town) who immediately starts work entering the data. The team moves on to the next selected village/cluster, Area Unit B. The team stays in Area Unit B for the 2nd week, and does the 1st interviews for another 16 households selected in that unit. At the end of this 2nd week, the team takes the questionnaires back to the data entry operator and

picks up the questionnaires for Area Unit A again. At the same time they will pick up the list of errors that they have made in those questionnaires and which may need to be checked in Area Unit A. They take these questionnaires with them to Area Unit A where, during the 3rd week, they undertake the 2nd interviews for the same 16 households, also checking the errors where necessary in their 1st week's work. At the end of the 3rd week they return to the data entry operator with the completed questionnaires for Area Unit A, and pick up the questionnaires, together with the list of errors, for Area Unit B. Then they go to Area Unit B again for the 4th week, which completes the cycle. At the end of 4 weeks they have completed 16 households in each of 2 area units.

In Option 1, the survey is conducted by 8 teams working for 10 periods of 4 weeks: sample size = 8 x 10 x 2 x 16 = 2560. In Option 2, there are 15 teams: sample size = 15 x 10 x 2x16 = 4800.

2.4.2 Number of households selected per village/cluster

It is clear from the above arrangements that the number of households selected in each village/cluster should be a multiple of 16. Should it be 16, or 32, or 48...?

This number is called the sample "take" in the area unit, and is often denoted by the symbol **b** . The problem of optimum take is discussed in sampling textbooks. Briefly, one computes the optimum value of b by making two simple mathematical models. Model 1 relates b to the sampling error; Model 2 relates b the to the unit operational costs. Smaller values of b lead to smaller sampling error but to higher cost, for a given total sample size. The problem of optimization is to balance these opposite trends, so as to obtain minimum error per unit cost. The optimum depends on the survey variable studied. In a survey of the present kind the optimum is likely to be around 20 households for some of the more important variables, based on experience in other countries.

It is recommended that the figure of 16 households be accepted for this survey as the number to be interviewed in each village/cluster.

However, one important reservation must be made at this point. There is always a possibility that a selected household may not be available, due to absence, illness, etc. This will be rare, but it will sometimes happen. The best way to deal with this problem is to select a small number of "spares" in each village/cluster, which can be used as replacements. It is recommended to select a total of 20 households in each village/cluster, of which 4 will be designated for use only where necessary to replace a household that cannot be interviewed. The number to be interviewed should always be 16.

2.4.3 Number of villages/clusters selected per commune

Transportation in Viet Nam is difficult and costly. The distances traveled can be reduced by grouping the villages in pairs since the field arrangements require the team to travel twice between each pair of villages. In the urban sector, there will be a saving if the two clusters are in the same town; one way of ensuring this is to select them in the same commune. Thus there is some advantage in selecting an even number of villages/clusters in each commune.

Note, however, that the larger the number of villages/clusters selected per commune, the smaller the number of communes in the sample. A smaller sample of communes reduces the amount of sampling work, as well as the amount of travel. However, it will certainly increase the sampling error: for sampling efficiency the main consideration is to have a widespread 1st stage sample. If 4 villages/clusters were selected per commune the number of communes selected in Option 1 would be 75 (of which 15 urban), and in Option 2 it would be 40 (of which 13 urban). These numbers are certainly too small for a satisfactory sample.

It is therefore recommended to select just two villages/clusters per commune.

2.4.4 Number of communes to be selected

These numbers already emerge from the above discussion. With 16 households to be interviewed per village/cluster and 2 villages/clusters selected per commune, the number of clusters in the sample will be: for Option 1, 150, and for Option 2, 80.

3. STRATIFICATION

3.1 Purpose

The purpose of stratification is to reduce sampling error. This is achieved by dividing the sampling frame into sections, or strata, and carrying out the sampling independently in each stratum. In this way the number of sampling units selected in each stratum is fixed by the sample designer. If stratification is not used these numbers are determined by chance. If the strata are very different from one another this chance may make a large contribution to the sampling error. To take an extreme example, it could happen, with strictly random selection, that all the communes selected are in the south of the country; with stratification this could not occur.

The gain from stratification depends on creating strata that are as internally homogeneous as possible. The gain comes mainly from stratification at the 1st sampling stage: communes in the present case.

In the case of Viet Nam the available area sampling frame appears to contain no data relevant to stratification except the geographical location of the units and their classification as urban or rural. It is therefore recommended that the urban and rural sectors be the two principal strata, and within these a geographical stratification be applied. The method proposed is given below.

3.2 Systematic sampling

Systematic sampling means sampling at a fixed interval from a list, beginning at a random starting point. This procedure ensures that the sample is well spread out through the list. Since most lists are arranged in some kind of systematic order, at least roughly, the effect is similar to stratification. Such sampling is also easier than using random numbers to obtain a truly random sample. Thus in practice systematic selection is much more often used than random selection. Its use in the present survey is recommended, both for selecting areas and households.

In the case of the 1st stage, selection of communes, the selection should be made from a list of communes arranged by province and the provinces should first be arranged in geographical order from North to South. Since economic variables tend to reflect agricultural factors, and these reflect the nature of the terrain, it is desirable to prepare such a list in order of geographical continuity. That is to say, the provinces should be listed so that two provinces which are neighboring in the list will be neighboring on the ground. This is most easily done by listing the provinces in "serpentine order".

Two lists are needed, one of rural communes, classified by province with the provinces listed in serpentine order; the other of urban communes, again classified by province with the provinces listed in the same order again. The selection of communes is then carried out independently in the two lists.

4. THE PPS SELF-WEIGHTING SAMPLE DESIGN

This section describes a well known sample design in which the 1st stage units are selected with probability proportional to size (PPS) and a fixed take is used at the final stage, the overall probabilities being constant for all final units (households). The method has several advantages and is recommended for the present

survey. In Section 4.1 the method is described in its simplest form, the case of 2-stage sampling. In Section 4.2 the method is extended to 3-stage sampling and described in terms of the present survey. Section 4.3 gives the precise procedures used for sample selection with PPS.

4.1 The 2-stage case

Two stage sampling is assumed, with area units at the 1st stage and households at the 2nd.

The 1st-stage units (primary sampling units, or PSUs) are selected with probability proportional to size (PPS). It is assumed that the number of households in each PSU is known: this is the "size" of the PSU and may be denoted by M_i for the i-th PSU. Then the probability of including this unit in the sample is:

$$p_{1i} = k M_i \qquad \text{---------- (1)}$$

where k is a constant. The subscript 1 is used to denote the 1st stage of sampling and i denotes the i-th PSU. It is easily shown that the constant k is equal to $a/\Sigma M_i$, where a is the number of PSUs selected and ΣM_i is the sum of M_i over the whole sampling frame.

At the 2nd stage a fixed number b of households is selected in each PSU from the M_i existing. This implies a 2nd stage probability:

$$p_{2i} = b/M_i \qquad \text{---------- (2)}$$

This is the conditional selection probability for any household in the i-th PSU, that is, the probability of selecting the household given that the PSU i has already been selected.

The overall probability of selection F for any household is the product of p_1 and p_2 , that is, the probability of first selecting the area multiplied by the probability of then selecting a given household. Thus:

$$F = p_{1i}\, p_{2i} = (aM_i/\Sigma M_i).(b/M_i)$$

$$= ab/\Sigma M_i \qquad \text{---------- (3)}$$

This is equal to the ratio of the sample size to the population size (number of households) - which was to be expected since the overall probability is simply the sampling fraction.

The important feature of this design is that, as equation (3) shows, the overall probability for households, F, is constant. Thus all households have the same chance of selection and the sample is "self-weighting". Note that the assumption is made that

the number of households M_i is the same in equations (1) and (2). In practical applications this is not necessarily true: the value in equation (1) comes normally from the census, while the value in equation (2) comes from a listing operation made just before the survey. (However, if the two values are not equal but in an equal ratio in all PSUs, self-weighting will still apply.)

In the absence of any information about differences between areas as regards population variances and operational costs, such a sample is likely to be of approximately optimal efficiency, that is, of minimal sampling error per unit of operational cost. This assertion assumes that the over-riding objective is the provision of national estimates. If greater priority is given to reducing sampling error at the regional (sub-national) level this conclusion does not hold. In the present survey it is thought that priority should be given to the national estimates.

In summary, the design has three advantages:

1) Approximately optimal efficiency for national estimates;

2) A fixed "take" of households in each area unit - which is an organizational advantage;

3) A self-weighting sample, which simplifies analytic work and record-keeping.

4.2 Three-stage sampling: the present survey

The same method extends in a very simple way to any number of stages. Each stage except the last is selected with PPS, and in each stage the "take" is fixed.

Assuming 3 stages, as in the present survey, the relevant equations are:

1st stage: $p_{1i} = am_i/\Sigma M_i$

2nd stage: $p_{2ij} = bM_{ij}/\Sigma_j M_{ij}$

3rd stage: $p_{3ij} = c/M_{ij}$

When the three probabilities are multiplied together, the overall probability is now:

$$F = abc/\Sigma M_i \qquad \text{---------- (4)}$$

In the present survey, c = 16 and b = 2.

As before, we have to take note of certain assumptions that are made here. It is assumed that the M-values are consistent.

That is, we assume that the total across all villages in a commune, as reported by the commune for sampling stage 2, is equal to the census total for the commune as used for selection in stage 1. In symbols, $\Sigma_j M_{ij} = M_i$. We also assume that the total number of households listed for the 3rd stage in any village/cluster ij is equal to the number M_{ij} reported by the commune for stage 2. (Once again, if these numbers are not equal but in an equal ratio in all area units, self-weighting will still apply.)

4.3 Method of sample selection with PPS

This section describes a well known method for systematic selection with PPS which is recommended for use at the 1st and 2nd sampling stages in the present survey.

The method is first described as it applies to the 1st stage: selection of communes.

List the communes in a column, with the measure of size M_i entered opposite each one. Cumulate these M_i in a further column headed cum M_i. Obtain the sampling "step" S by dividing the total M_i (= last figure in the cum M_i column) by the number of communes wanted in the sample. Thus, $S = \Sigma M_i/a$. Obtain a random number C, less than or equal to S. Compute the sampling sequence:

$$C \; ; \quad C+S \; ; \quad C+2S; \quad C+3S \; \ldots$$

For each term of this sequence the unit selected is the first one whose cum M_i equals or exceeds that term.

In the case of the 2nd stage sampling (selection of villages or clusters), the method is the same. A separate step has to be computed for each commune i, using $S_i = \Sigma_j M_{ij}/b$. In the present survey b = 2. A separate random number C_i and a separate sampling sequence will also be required for each selected commune i.

The 1st stage selection should be carried out separately within each of the explicit sample strata, that is the rural and urban sectors.

5. PARAMETERS FOR THE 1ST STAGE SAMPLE

The allocation of the sample between the rural and urban domains was discussed in Section 2.2. It is now appropriate to give the precise parameters.

The first step is to obtain the census distribution of households between the rural and urban sectors. The data below take account of the requirement mentioned in Section 2.3.2 above:

individuals living in collective households are considered as 1-person households. Thus the figures show the number of family households plus the number of persons in collective households.

Table 1 below shows this breakdown, together with the sampling fractions and sample sizes, separately for Options 1 and 2.

TABLE 1 PARAMETERS FOR THE 1ST STAGE SAMPLE

	R U R A L	U R B A N	T O T A L
Number of households in census	N_R = 10440156	N_U = 2479501	N = 12919657
OPTION 1			
Sampling rate	F	F	
Sample size	FN_R	FN_U	4800
OPTION 2			
Sampling rate	F_R	F_U	
Sample size	F_RN_R	F_UN_U	2560

From these data the necessary sampling step for selection of commnunes can be computed, for the rural and urban strata and for Options 1 and 2.

Option 1

Since 32 households are to be interviewed in each commune, the number of communes required in the sample is 4800/32 = 150. Dividing this number between rural and urban in proportion to the values N_R and N_U, yields:

Rural: 121 communes. Urban: 29 communes.

These are the values called a in equations 3 and 4. Dividing them into the corresponding values of N one obtains the sampling steps:

Rural: 86282 Urban: 85500

Option 2

Here the urban sampling rate has to be made equal to twice the

rural, or $F_U = 2F_R$. Substituting this in the last line of Table 1 and summing across, one obtains:

$$F_R N_R + 2F_R N_U = 2560,$$

so that $$1/F_R = (N_R + 2N_U)/2560 = 6015.3$$

Using this step would yield 10440156/6015.3 = 1735.6 households in the rural sample. Allowing for 32 households selected per commune, this amounts to 54.24 communes. This has to be rounded to exactly 54 rural communes. Since the total number of communes is 2560/32 = 80, this implies 26 urban communes. Returning to apply these numbers to the values of N_R and N_U, one obtains the sampling steps:

Rural: 193336 Urban: 95365.

NOTE ON SAMPLE EXECUTION

At the time of writing these parameters have been used in the selection of a sample of 80 communes for Option 2. In the case of Option 1 some minor errors were found only after much of the work had been done. Rather than repeat the work some minor adjustments were made. The resulting sample of communes corresponds to the correct number of urban and rural selections, and differs only trivially from the specification.

6. SAMPLING OF HOUSEHOLDS

6.1 Procedure for selection

Sampling of households within villages/clusters will be by systematic selection with equal probability. The procedure recommended is as follows.

The number of households to be selected is always 20. The first step is to obtain the lists of households in the selected village/cluster. Number these households from 1 upwards. Let the number of households listed be N_{ij}. First check whether this number agrees reasonably well with the number supplied by the commune during the initial request for data: if not there may be some misunderstanding.

Divide N_{ij} by 20 to obtain the step S_{ij}. Select a random number less than or equal to the step, = C_{ij}. Form the sampling sequence:

$$C \; ; \quad C+S \; ; \quad C+2S \; ; \quad C+3S; \; \ldots$$

Each term corresponds directly to the household bearing the same number and selects that household.

One problem requires special attention. At the household selection stage the sampling step may often by quite small, for example less than 5. If the step is rounded to the nearest whole number, the error may be substantial. It is better to compute the step to 1 decimal place. How is selection done with a decimal step? The procedure is as follows.

1) Express the step S rounded to one decimal place.

2) Find a random number C between 1 and 10S. Place a decimal point before its last digit. This becomes C.

3) Compute the sampling sequence:

C ; C+S ; C+2S ; C+3S ; ...

4) The whole number part of each sampling number indicates the unit selected.

EXAMPLE

Suppose the step S = 3.4 .

Select a random number between 1 and 34. Suppose 23 is selected. Then C = 2.3 . The sampling numbers and selections are as follows:

Sampling number	Unit selected
2.3	2
5.7	5
9.1	9
12.5	12
etc.	etc.

6.2 Selecting the spares

Among the 20 households selected, 4 are to be allocated for use as replacements, or "spares". It is recommended that these 4 be spread systematically through the list of 20 to avoid any systematic bias in the use of spares. A very simple but adequate scheme would be to allocate always the 3rd, 8th, 13th and 18th selections to be spares.

A separate list should be made of these spares, which should be made available to the field supervisor but not the interviewer. Interviewers should be allowed to use a replacement only if the supervisor authorizes it. The supervisor will always provide the

first spare from the list.

7. ALLOCATION OF THE SAMPLE TO COMMUNES AND MONTHS

When the sample of communes has been finalized (and this means when a decision has been made between Options 1 and 2), the next step is to work out a detailed time-table of operations, showing the travel of the teams, so that the date on which each commune is visited will be exactly known.

In principle it is desirable that the sample be spread over the whole country at all times, so that team movements do not interact with seasonal changes. In practice this may not be perfectly feasible because of climatic constraints on access to certain communes at certain seasons. Some compromises will have to be accepted but the guiding principle should be followed as far as possible.

8. ESTIMATION AND SAMPLE WEIGHTING

At the end of Section 4.2 it was pointed out that the self-weighting assumption requires that the various household counts are mutually consistent, or at least that they retain a fixed ratio to one another throughout the sample. How far can this assumption be accepted in the present case?

The census was carried out in 1989 and the survey will be implemented in 1992. During this interval the population will have increased by something between 5 and 10%. The increase in the number of households should be rather smaller. If these changes are spread uniformly over relevant population groups (such as urban/rural, North/South, rich/poor) self-weighting will be maintained.

Natural population growth is likely to be fairly uniform. The other main changes of importance have been the return of workers from Irak and the return of refugees from Hong Kong and elsewhere. While these have involved substantial numbers, in a high proportion of cases these individuals will have returned to existing households, rather than forming new households. Since the present concern is with numbers of households, the effect of these movements seems likely to be small.

Note finally that the Government operates strict control on movement into the large cities, and to a lesser extent on all migratory movement.

In addition to these substantive factors, it should be borne

in mind that, with demographic and socio-economic variables, small changes in weights of broad groups have very small effects on estimates.

In my view it is justifiable in the light of the above considerations to regard this sample as self-weighting

→ INSERT attached

ANNEX. FIELD VISITS

Visits were made to two communes on 27 November 1991, one urban in Hanoi, the other semi-rural.

Visit to Truc Bach

This is a commune in central Hanoi with a census population of 11,876.

The Commune Head reported a current population of 12,518, or 3015 households. We were shown lists of Cum by population. We visited one Cum and were shown a household list. The authorities concerned were extremely cooperative and said that there would be no difficulty in providing the information which would be required. All the lists are up-dated at least every 6 months.

Visit to Van Yen

This is a semi-rural commune on the outskirts of Hanoi in the Province of Ha Tay. The census reports 8614 population, or 2070 households. There is a substantial institutional population consisting of persons who work in Hanoi.

We were given a list of 4 Thon, with their farming households. The authorities had some difficulty in giving the same information for the state employees but said that there would be no difficulty given time.

We were well received everywhere and formed a very positive impression about the feasibility of the sampling arrangements that had been proposed.

INSERT ON PAGE 16

The 2nd paragraph on page 16 should read as follows:

In my view it is justifiable in the light of the above considerations to regard this sample as self-weighting in the case of Option 1; in the case of Option 2 it is self-weighting within the urban sector, and separately within the rural sector, but not across the two together. In Option 2a weight of 2 should be applied to the rural sector relative to the urban. (Strictly, the required weight is 193336/95365, which is equal to 2.027 . The approximation of 2 exactly is fully acceptable.)

Thus in Option 2, estimates which include both rural and urban data should be based on the formulas:

For a mean Y/N, estimate $\sum y_i w_i \; / \sum n_i w_i$

For a ratio Y/X, estimate $\sum y_i w_i \; / \sum x_i w_i$

For a proportion p = N'/N, estimate $\sum n'_i w_i \; / \sum n_i w_i$

For a percentage P = 100N'/N, estimate $100 \sum n_i w_i \; / \sum n_i w_i$